2013

GO GIRLS GO Cyber : Data Miner

Department of Computer Science

Wayne State University

Contents

Chapter 1 – Getting Started 3
Overview 3
About the User Manual 3
Installation Guide 3
Prerequisites: 3
Set Up 3
Chapter 2 – Usage 4
Chapter 3 – Recommender System 5
3.1. Interactive Part 5
3.2. Demonstration 7
Chapter 4 – Cluster Analysis 10
Chapter 5 – Classification 12
5.1. Beginner 12
5.2. Intermediate Level 13
5.3. Advanced Level 14
Chapter 6 – Association Analysis 17
Chapter 7 – PageRank 19
Chapter 8 – Social Networks 23
8.1. Beginner's Level 23
8.2. Intermediate Level 24
8.3. Advanced Level 25
Contact Us 27

Chapter 1 – Getting Started

Overview

As we are in the information age, we believe that information leads to power and success. Today, we have far more information than we can handle which leads to find new ways to help us make better managerial choices. Data mining, a powerful computing technology plays a major role in this situation which is an analysis of data to extract useful patterns or information from a large data collection.

It is important to educate the youth generation about these novel technologies and directs them for higher education in these novel areas. As Data Miner: Educational Tool was built for middle school children, its primary objective is to teach them the Data mining concepts in a simplified manner and their real world application, hence motivate them to follow their higher education in Computer Science.

About the User Manual

The user manual of Data Miner covers all the details of the tool including data mining concepts implemented.

Installation Guide

Prerequisites:

- Adobe Flash Player
- Web browser

Set Up

- Click on the link provided.
- Follow the instructions provided in this manual for each module.

Chapter 2 – Usage

Data Miner describes six main data mining concepts in separate modules,

- Recommender System
- Cluster Analysis
- Classification
- Association Analysis
- PageRank
- Social Networks

Data Mining
Hi... I'm Mynee, your Data Mining assistant. I'm here to guide you to work with this application. If you need any help, count on me !! First move your mouse over these buttons to see what you can do.
RECOMMEND TO ME
CLUSTERING
CLASSIFICATION
ASSOCIATION ANALYSIS
PAGERANK

Figure 1 Home Page

Each module consists of a welcome page that includes a brief description of the concept, demonstration and interactive part. In demonstration, it explains how to plays at the interactive part. In addition to that some phases include quizzes to evaluate the user.

Chapter 3 – Recommender System

Data Miner demonstrates Recommender System by movie preferences and explains two approaches such that recommendations based on the contents of the movies and recommendations based on the other users preferences.

3.1.Interactive Part

In the first approach, user will be provided two sets of movies such that each set contains ten movies. Any movie belongs to a single genre and the system considers only five genres; *Adventure, Animation, Comedy, Family and Fantasy*. This approach recommends items that are similar to those that a user liked in the past. User can LIKE movies and depending on the LIKEd movies, system will recommend some different movies to the user.

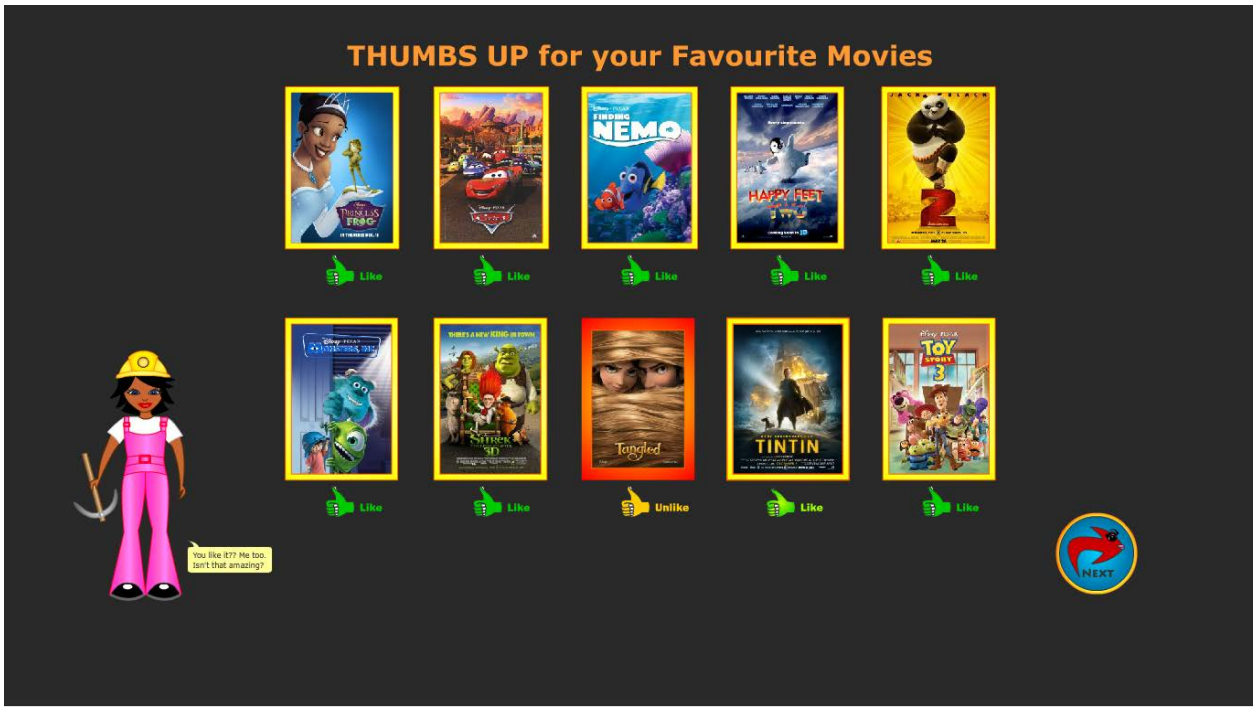

Figure 2 Recommender System Interactive Part 1 - Inputs

User can hit on *Like* button to like the movie and it becomes an *Unlike* button. He can click on the same *Unlike* button to undo the action, i.e. Unlike the liked movie. Once user finished his selection, he can move to the next step to see the recommendation by clicking on *Next* button. To make the Next button work user has to like at least a single movie. Next page shows the *Recommendation* button and when the user clicks on it, it will lead to the recommendation movies for the user's preference.

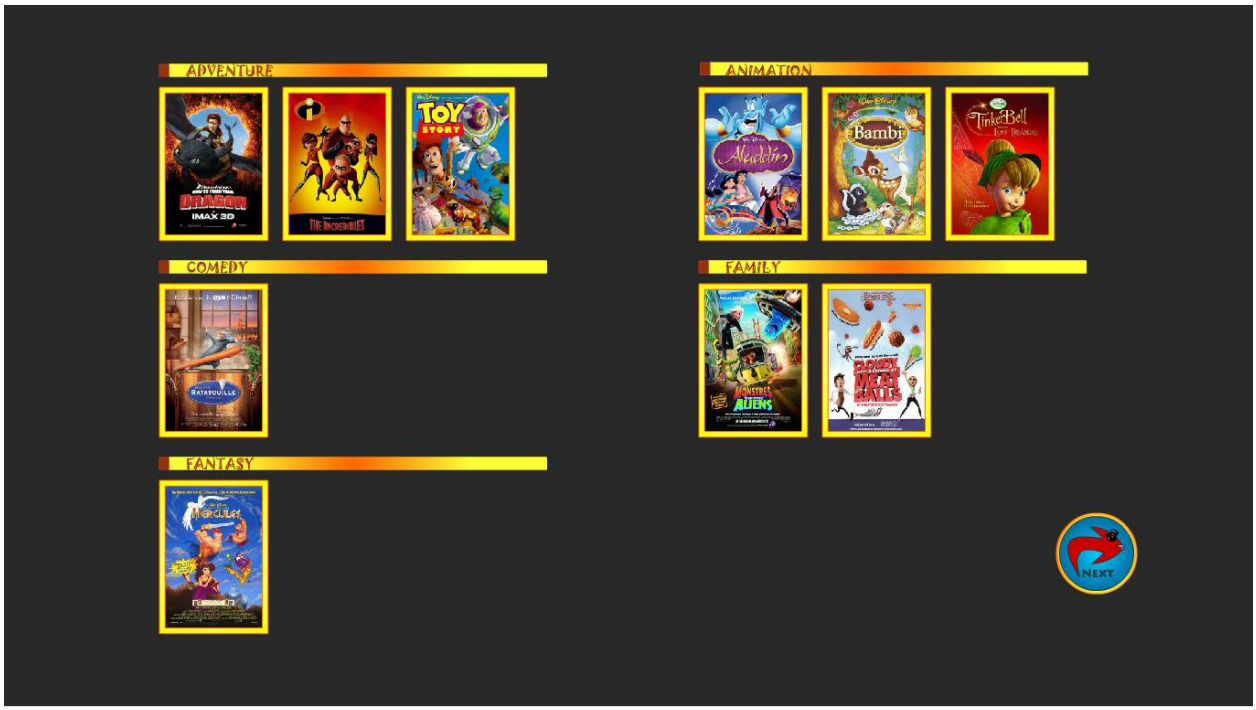

Figure 3 Recommender System Interactive Part 1 - Output

The second approach contains three quizzes and it teaches how to recommend movies based on friends preferences. The Next button and Back button will not work until the correct answer is given.

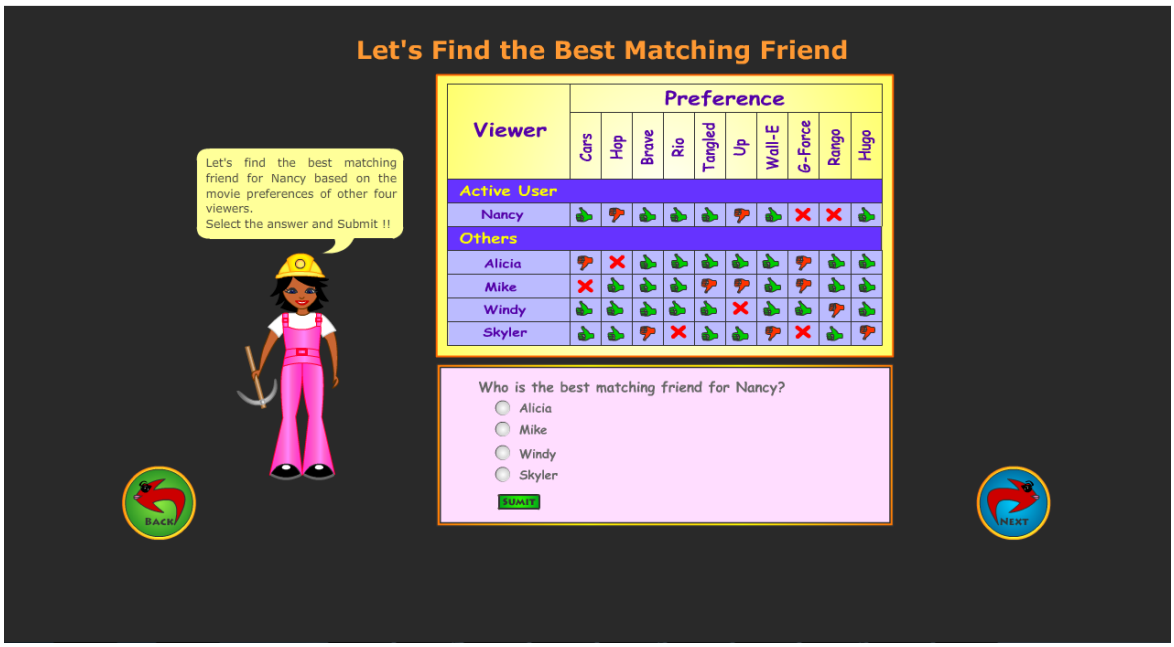

Figure 4 Recommender System Quiz 1

3.2. Demonstration

Recommender System includes a demonstration of Collaborative filtering which explains the underlying mechanism of finding the similar user. To see how it works, user has to click on *Step 1, Step 2*, and *Step 3* buttons in each page.

STEP 1:

Click on *Step 1* button to play the animation. Once it is completed hit *Next* button to see the next step.

How Recommender System Works?
STEP 1
COLLECT INPUT
Lets see how Recommender System suggests you a friend.
First click Step 1 to see how you entered the name and preference for each movie.
Click Next go to Step 2.
Name: Jaden Smith
THUMBS UP For Your Favourite Movies
BACK
NEXT

Figure 5 Demonstration Step 1

STEP 2:

Click on Step 2 button to play the animation. Click on *Next* button in the demonstration part (not the Circled *Next* button) to calculate similarity score for each user. When the step 2 demonstration is over, click on *Next* button outside to move to the step 3.

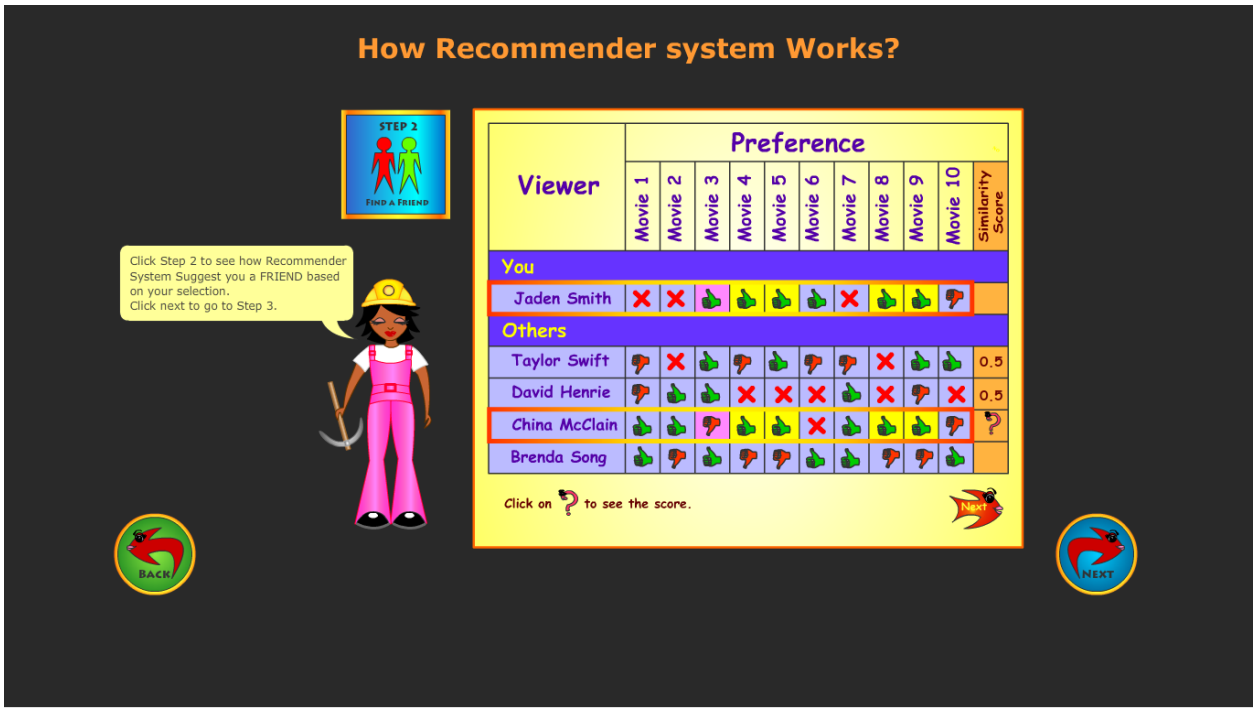

Figure 6 Demonstration Step 2

STEP 3:

Click on *Step 3* button to play the animation. Once it is completed hit *Next* button to go to the second movie set of the collaborative filtering.

When the last phase of collaborative filtering is completed, the Recommender System reaches to its end and clicking on *Next* button directs to the Home Page while clicking on Back button directs to the initial page of Recommender System.

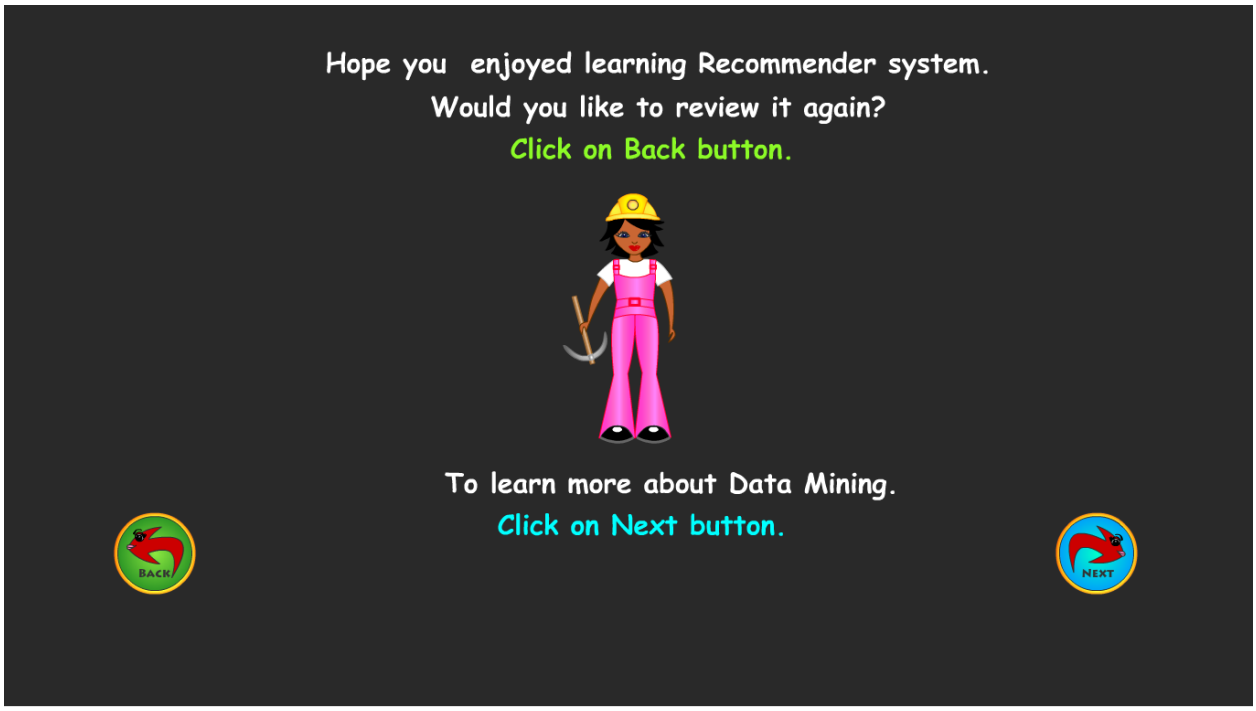
Hope you enjoyed learning Recommender system.
Would you like to review it again?
Click on Back button.
To learn more about Data Mining.
Click on Next button.
BACK
NEXT

Figure 7 Recommender System Last Page

Note each data mining concept ends with a page like this with similar behavior.

Chapter 4 – Cluster Analysis

Objective of Cluster Analysis is teaching how to group a set of objects based on their attributes such that object with similar attributes belong to the same group while dissimilar object belong to different groups. It contains a demonstration as an animation clip which shows how the interactive part works.

The interactive part of Cluster Analysis contains three levels such as Beginner's level, Intermediate level and Advanced level. Each level will gradually increase the difficulty of grouping given objects. The objects given to be clustered contain a list of attributes which are displayed for Mouse Over event. Since this module simulates the Cluster Analysis, clusters are predefined and the number of containers indicates the number of cluster to be created. User has to click on an object, drag and drop it into one of the cluster containers provided. Even though cluster labels are unknown at the beginning, as cluster objects of each container is predefined, once user dropped an object to a container, the object will stay with it if that object belongs to that cluster. If it's not the right container, the object will move to its initial position. The assistant of the tool will provide a hint about the clusters to make the task easier for the user.

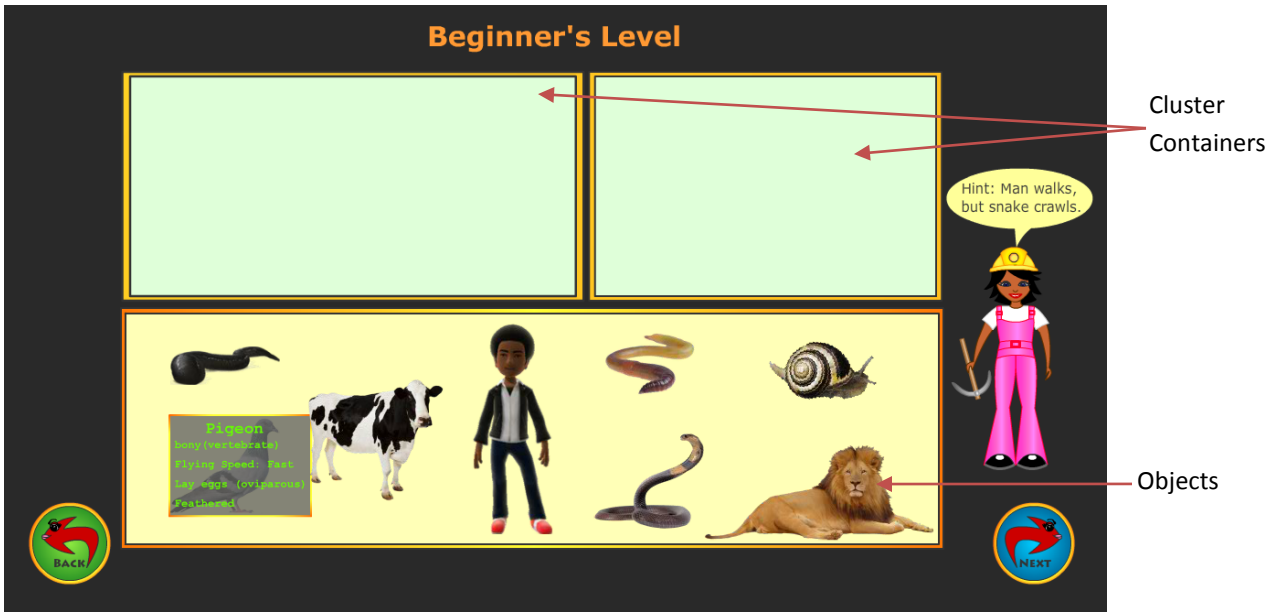

Figure 8 Cluster Analysis Interactive Part (Beginner's Level)

Note that objects should drop in to the middle of the container such that it does not overlap with edges of the container. At the end of each level, a quiz will be given to evaluate the user's understanding about attributes based on clustering. The system indicates whether the submitted answer is right or wrong and does not allow moving to the next level until, the right answer is submitted.

Next level can be reached by clicking on the *Next* button and *Back* button directs to the Difficulty Level Selection page.

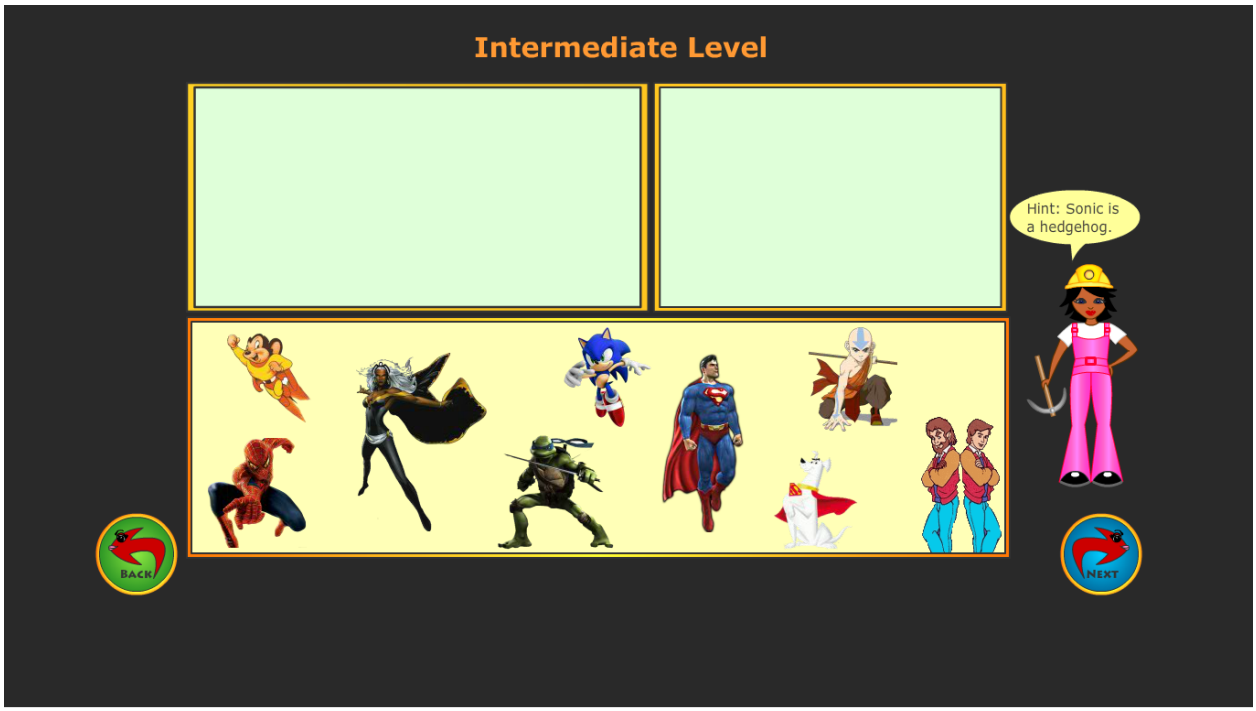

Figure 9 Cluster Analysis Intermediate Level

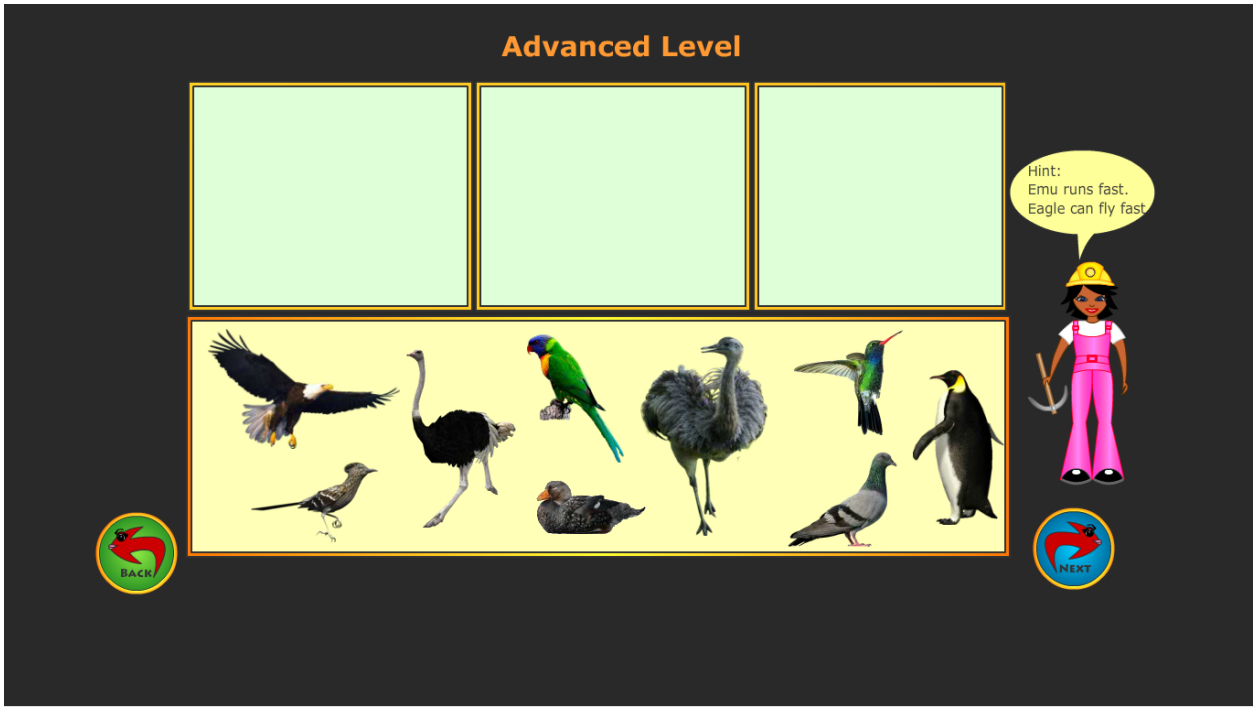

Figure 10 Cluster Analysis Advanced Level

Chapter 5 – Classification

Basic idea behind Classification is assigning given items to a known class which have items with similar attributes. In Data Miner, Classification is explained by building decision trees by dividing set of items by values of a specific attribute in each level and selecting the best split such that each classification is pure, i.e. no branch contains items which belong to different classes. This phase contains a demonstration of classification using decision tree as an animation clip which shows how the interactive part works.

Similar to the Cluster Analysis, the interactive part of Classification contains three levels such as Beginner's level, Intermediate level and Advanced level. Each level will gradually increase the difficulty of classifying given objects by increasing the number of attributes, classes and levels involve with the tree. The objects given to be classified contain a attributes and attribute values which are displayed for Mouse Over event. To make the tool less complicated, the number of branches and nodes (containers) for each branch of the tree are given to the users, so they can drag and drop items to these containers just like the objects behave in Clustering phase.

Figure 3 Attributes, Attribute Values and Class label

5.1.Beginner

Beginner's level consists of ten objects (animals) which belong to either one of two classes called *Mammals* and *Fish*. It provides two attributes *Give Birth* and *Live in Water* to split the root node. As each attribute can have either of two values *Yes* and *No*, for the selection of each attribute, two branches will be popped up. Beginner's level concerns only a single level of the tree which means it only allows to split the tree once. The objective of this level is to teach building a decision tree with a single level from the given set of items. Following is the table which contains animals listed in this level.

Name	Give Birth	Live in Water	Class
Platypus	No	No	Mammals
Human	Yes	No	Mammals
Cat	Yes	No	Mammals
Porcupine	Yes	No	Mammals
Bat	Yes	No	Mammals
Salmon	No	Yes	Fishes
Eel	No	Yes	Fishes
Leopard Shark	Yes	Yes	Fishes
Whale	Yes	Yes	Mammals
Dolphin	Yes	Yes	Mammals

Table 1 Classification Beginner's Level Objects

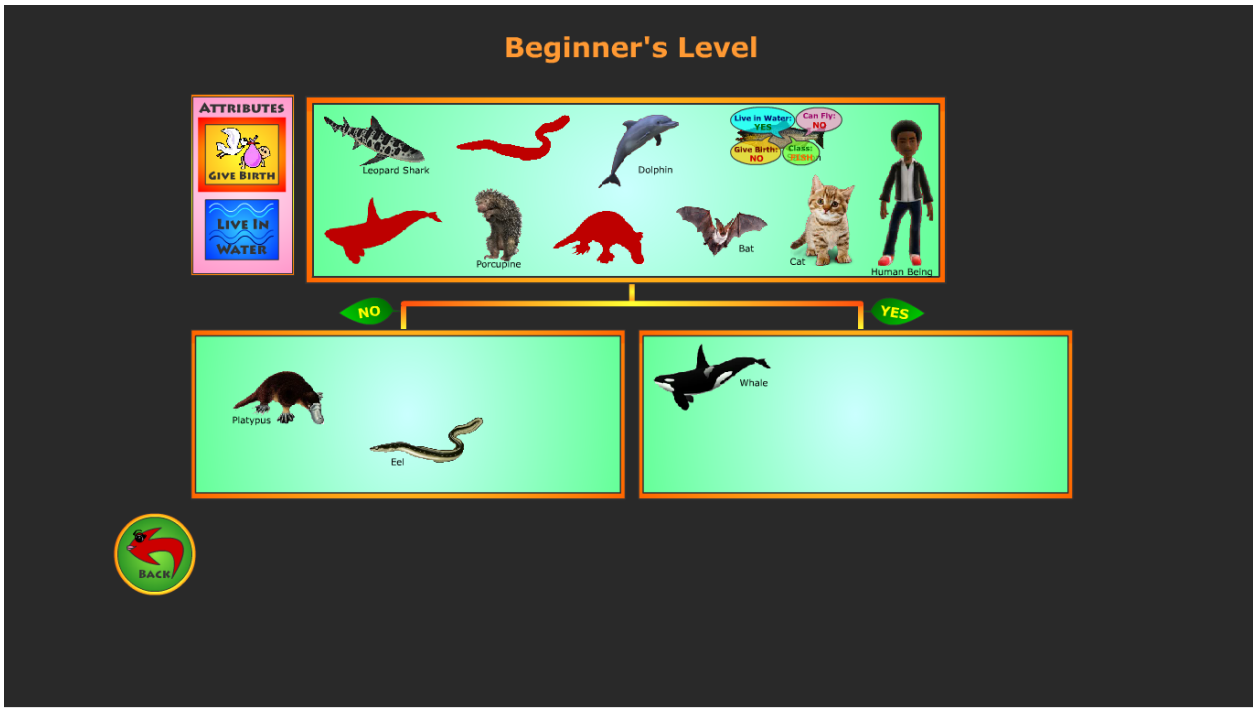

Figure 4 Classification Beginner's Level

For an instance, when user selects the attribute *Give Birth*, the root not is split into two branches with attribute values *Yes* and *No*. User can drag and drop items from root node, and if the animal is dropped into the right branch, it will stay with that node. Otherwise it returns back to the initial position in root node.

As this level provides two attributes to split the root, it can build two different trees. But neither split will make a pure classification.

5.2.Intermediate Level

Intermediate level consists of ten animals which belong to either one of two classes called *Mammals* and *birds*. It provides three attributes *Give Birth, Live in Water* and *Can Fly* and each attribute can have either one of two values *Yes* or *No*. The Intermediate level also concerns only a single level of the tree. The objective of this level is to teach building a decision tree with a single level from the given set of items and the difference between purity and impurity of the classification.

Following is the table which contains animals listed in this level.

Name	Give Birth	Can Fly	Live in Water	Class
Platypus	No	No	No	Mammals
Human	Yes	No	No	Mammals
Cat	Yes	No	No	Mammals
Porcupine	Yes	No	No	Mammals
Pigeon	No	Yes	No	Birds
Owl	No	Yes	No	Birds
Eagle	No	Yes	No	Birds
bat	Yes	Yes	No	Mammals
Whale	Yes	No	Yes	Mammals
Dolphin	Yes	No	Yes	Mammals

Table 2 Classification Intermediate Level Objects

Intermediate level provides three attributes to split the root and it can build three different trees. Even though neither split will make a pure classification, each classification creates one branch pure.

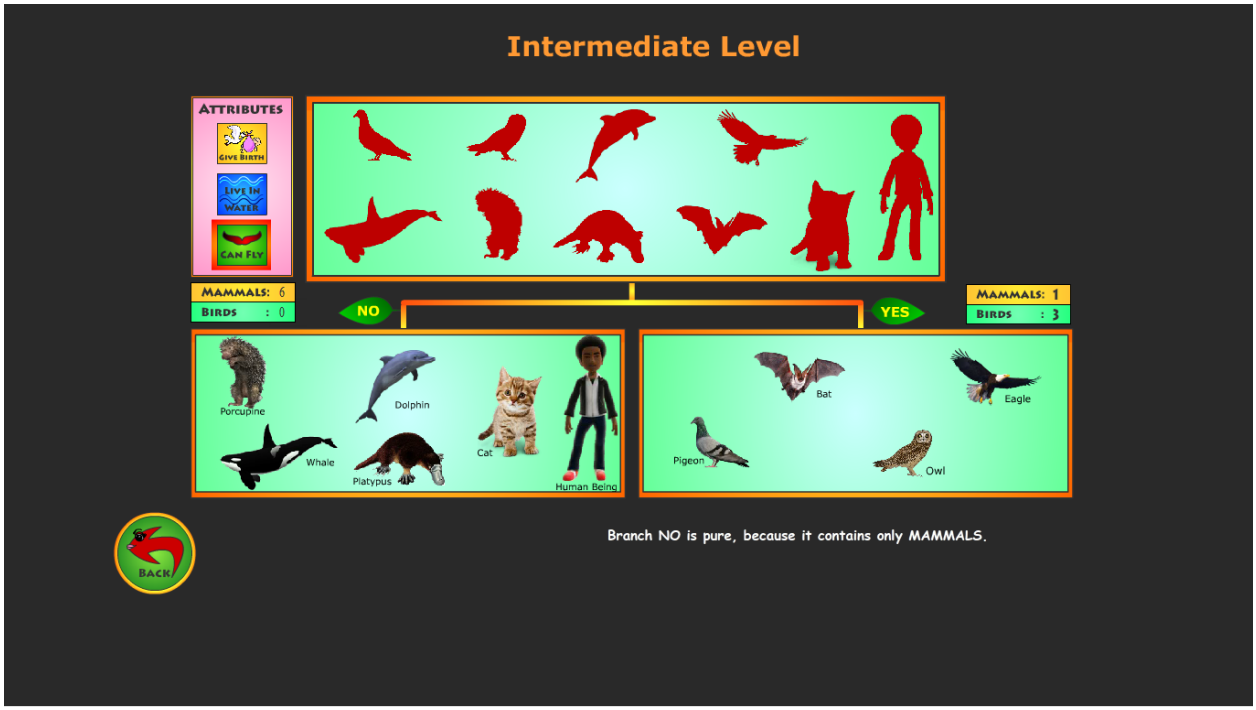

Figure 13 Classification Intermediate Level

5.3.Advanced Level

Advanced level also consists of ten animals which belong to either one of three classes called *Mammals, Fish* and *birds*. It provides three attributes *Give Birth, Live in Water* and *Can Fly* and each attribute can have either one of two values *Yes* or *No*. unlike other two difficulty levels, the advanced level concerns

two level of the tree to be split. The objective of this level is to teach building a decision tree from the given set of items with two levels until it build pure branches.

Following is the table which contains animals listed in Advanced level.

Name	Give Birth	Can Fly	Live in Water	Class
Pigeon	No	Yes	No	Birds
Owl	No	Yes	No	Birds
Eagle	No	Yes	No	Birds
bat	Yes	Yes	No	Mammals
Human	Yes	No	No	Mammals
Cat	Yes	No	No	Mammals
Porcupine	Yes	No	No	Mammals
Python	No	No	No	Reptiles
Komodo	No	No	No	Reptiles
Dolphin	Yes	No	Yes	Mammals

Table 3 Classification Advanced Level Objects

Intermediate level provides three attributes to build the tree into two levels; hence it can build six different trees. Unlike in other two levels, some splits will make a pure classification.

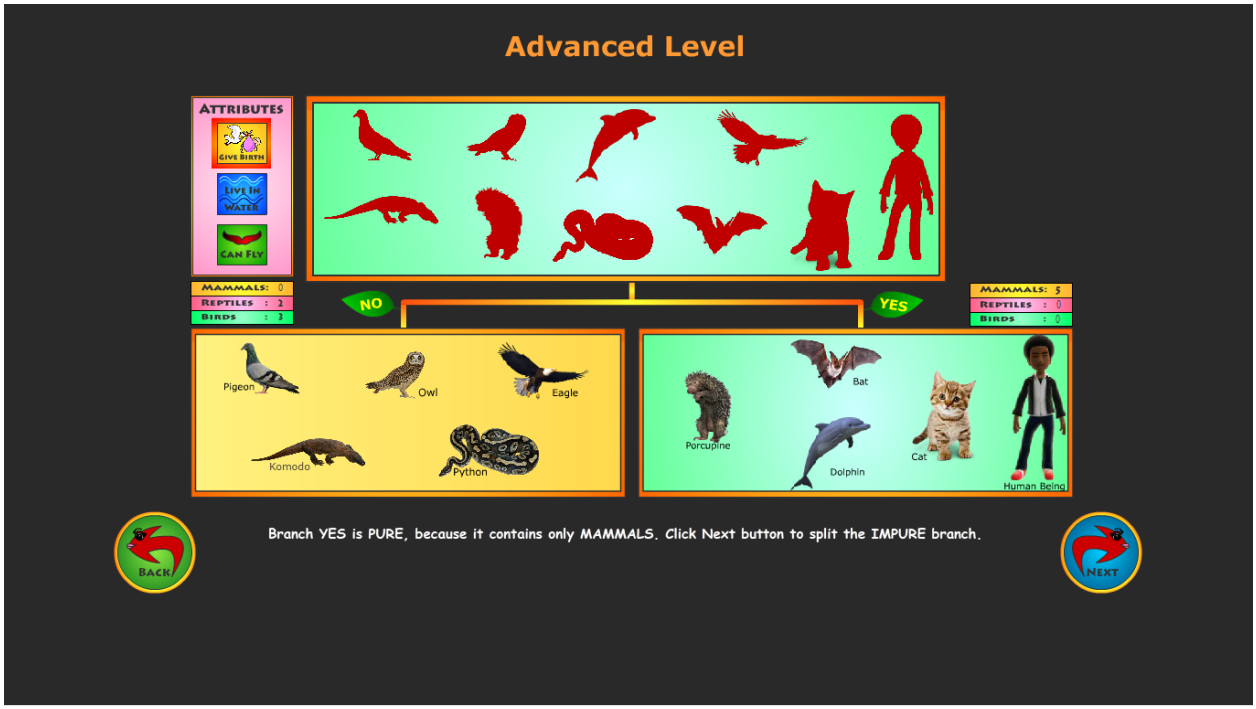

Figure 54 Classification Advanced Level

Figure 65 Classification Advanced Level - Pure Classification

Figure 76 Classification Advanced Level - Impure Classification

Chapter 6 – Association Analysis

This phase describes Association Analysis using Market Basket Transactions which is useful discovering interesting relationships in a set of transactions and talks about the Support Count which means the number of items present in a transaction. Similar to the other data mining techniques, Association Analysis phase also includes a demonstration part and interactive part.

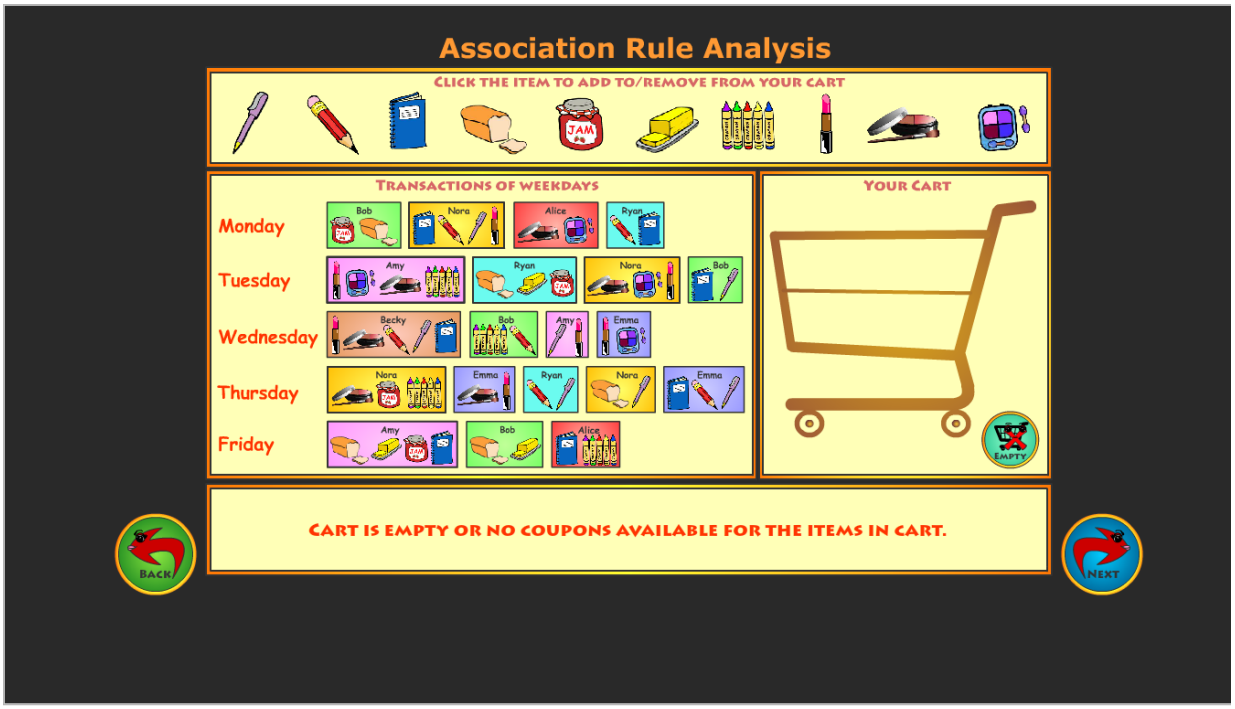

Figure 87 Association Analysis Interactive Part

Demonstration part explains how to work with interactive part and the idea behind that. In interactive part, the system provides a set of items available in a shop and user has to add them to his cart.

In addition to that, it provides a set of transactions occurred during five weekdays of a specific week such that each items is included at least in one of the given transactions. User can add an item to the cart by clicking on it and remove the same item from the cart by clicking on it again. When user adds an item to the cart, all the transactions which contains the item set in the cart will be highlighted (Figure 18). Depending on the associated items (common items) in these transactions which are not yet added to the cart, the system will display associated items about which item to be selected next in the area that associated items display with a coupon that offers some discounts depending on the items added to the cart.

Figure 98 Association Analysis – Content of Interactive Part

Once all the associated items are added to the cart, the system will indicate that to the user (Figure 19).

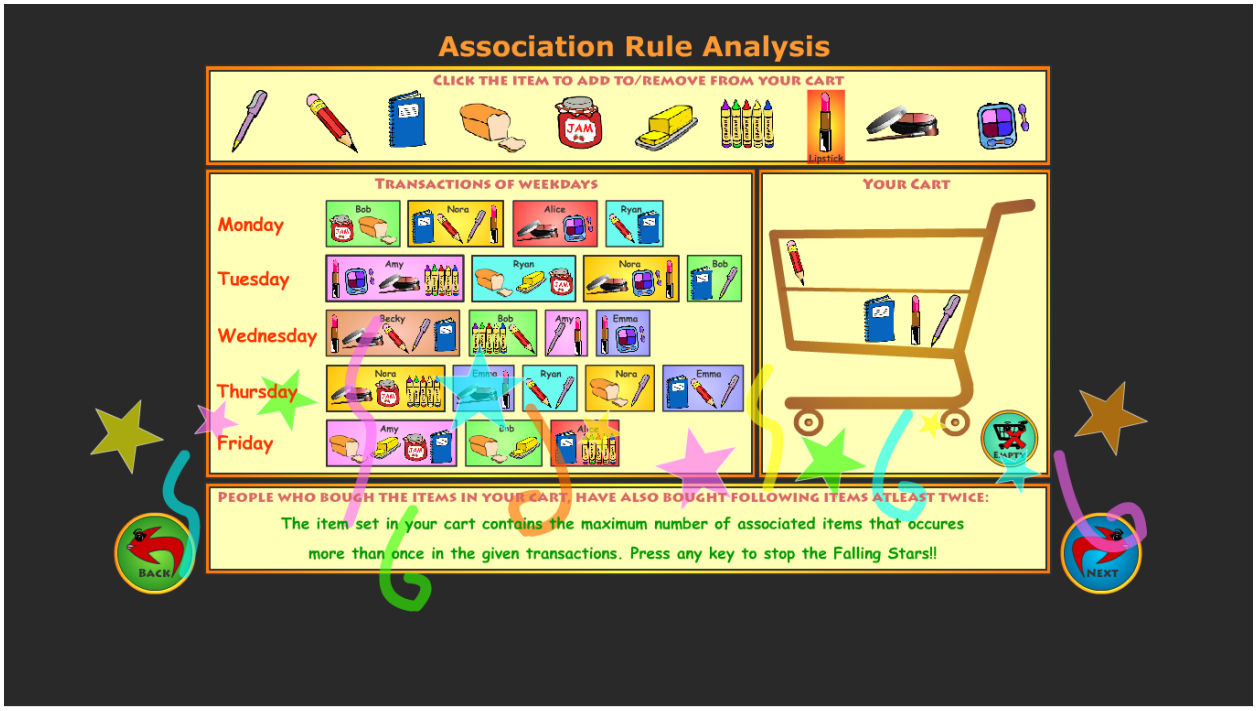
Association Rule Analysis
CLICK THE ITEM TO ADD TO/REMOVE FROM YOUR CART
TRANSACTIONS OF WEEKDAYS
YOUR CART
Monday
Tuesday
Wednesday
Thursday
Friday
PEOPLE WHO BOUGH THE ITEMS IN YOUR CART HAVE ALSO BOUGHT FOLLOWING ITEMS ATLEAST TWICE:
The item set in your cart contains the maximum number of associated items that occures more than once in the given transactions. Press any key to stop the Falling Stars!!
BACK
NEXT

Figure 19 Association Analysis - Completion of adding associated Items

Chapter 7 – PageRank

PageRank is a link analysis algorithm of the Google web search engine. It assigns a numerical weighting to each element of a hyperlinked set of documents, to measure relative importance of that element within the set. Objective of PageRank phase in this tool is to describe what Google PageRank is and how it affects to a certain web page or web site.

The demonstration of PageRank phase consists of two parts. The first part illustrates how to search over the Google search engine and what outcome will gain as results of search keywords. The second part describes the incoming and outgoing link of a web page.

Interactive part consists of two sections that demonstrate the same idea which is incoming links from popular, relative web sites increases the PageRank more than the links coming from non-relative, less popular pages. To explain this idea, the system uses a web site called “Go Music” which contains news related to music such as latest music albums released, popular artists, songs etc. User can link eight real web sites; Amazon, Danskin, Billboard, Pizza Hut, Yahoo, You Tube, AOL and Crayola to Go Music web site and see how PageRank of Go Music varies. Among these incoming links, Amazon, Billboard, Yahoo, You Tube, and AOL are considered as more popular than the other three web sites and contain news related to music. So links coming from these five web sites will show a higher increase in GO Music PageRank comparison to the links coming from other three web sites; Danskin, Pizza Hut, and Crayola.

The first page of interactive part describes the effect of incoming links using a High Striker. When a popular web page is linked to the Go Music web site, lever of the High Striker will hit the bell. If the incoming link is from a less popular page, then the lever will rise high, but will not hit the bell.

Features to Improve the PageRank
Link web sites to GO MUSIC
shelfari
amazon.com
DANSKIN
YAHOO!
Billboard
WIKIPEDIA
YouTube
AOL
GO MUSIC
Teen Choice Awards
Young people ages 13 to 19 can vote on their favourite actors, fashionistas, sports athletes and singers for the Teen Choice Awards.
This event, airing on the FOX Network in the United States, takes place every year at the Gibson Amphitheatre in California's Universal City.
Teenagers vote online to determine the winners for each of the categories.
Justin Bieber
Winners receive genuine custom made surfboards, each costing over $800. The surfboard represents the freedom teenagers experience during summer vacations.
New Music Releases
Red is the fourth studio album by Taylor Swift which was released on October 22, 2012.
"Take Me Home" is a record destined for commercial success.
"Call Me Maybe" is a song recorded by Canadian singer-songwriter Carly Rae Jepsen.
"Call Me Maybe" has attained commercial success worldwide, reaching #1 in many countries inclusing Australia, UK and US.
Watch Videos
The Gummy Bear Song
Mean - Taylor Swift
Skyscraper - Demi Lovato
Never Say Never
Related Links
More Information:
Teen Choice Awards
One Direction
Demi Lovato
SHOW MY WEB SITE
BACK
NEXT
If page rank increases largely, the lever of high striker will hit the bell.

Figure 10 PageRank Interactive Part 1

In the second page of the interactive part, Go Music is the center node and the user can link other nodes to it by clicking on them. To unlink a linked node, click on the same node. Green node represents popular and relative web sites, while Orange nodes represent the less popular web sites with non relative contents.

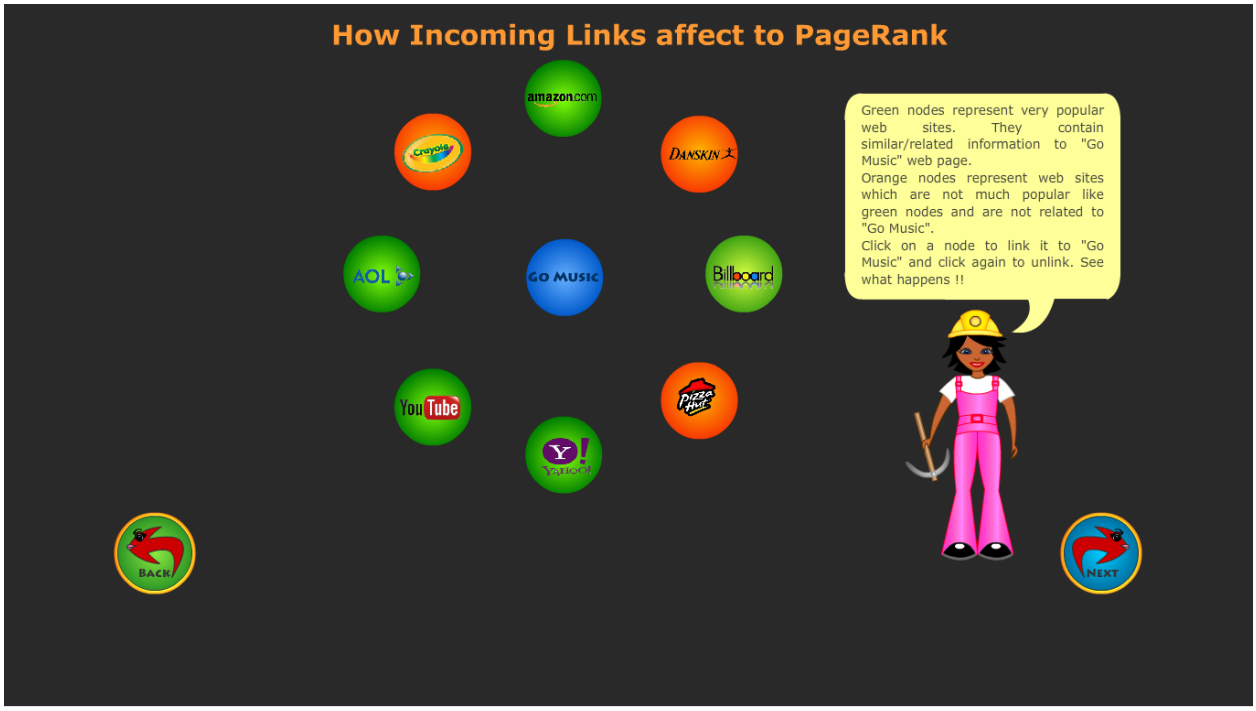
How Incoming Links affect to PageRank
amazon.com
Crayola
Danskin
AOL
Go Music
Billboard
YouTube
Pizza Hut
Yahoo!
BACK
NEXT
Green nodes represent very popular web sites. They contain similar/related information to "Go Music" web page.
Orange nodes represent web sites which are not much popular like green nodes and are not related to "Go Music".
Click on a node to link it to "Go Music" and click again to unlink. See what happens !!

Figure 11 PageRank Interactive Part 2

As the size of the Go Music node represents its PageRank, when a green node is linked to the Go Music, it becomes larger and unlinking the same green node changes Go Music node size to previous state. Linking Orange colored nodes do not change the size of Go Music.

(a) (b) (c)

Figure 12 Page Rank - Changes of Go Music Node depending on Incoming Links

(a) Initial Status
(b) Incoming Links fro less popular web sites
(c) Incoming links from popular, related web sites

Above three figures illustrates how center node size changes. In Figure 22 (a) shows the initial status of the Go Music node. According to Figure 22 (b), linking Orange colored node does not affect to the GO music node size. In Figure 22 (c), linking Green colored nodes changed the size of the center node. At the end of these two sections, user will be evaluated by a quiz.

The last interactive part of PageRank module describes the effect of incoming links from related web sites on Google search results. First user has to click on Play button to see the initial Google search results page.

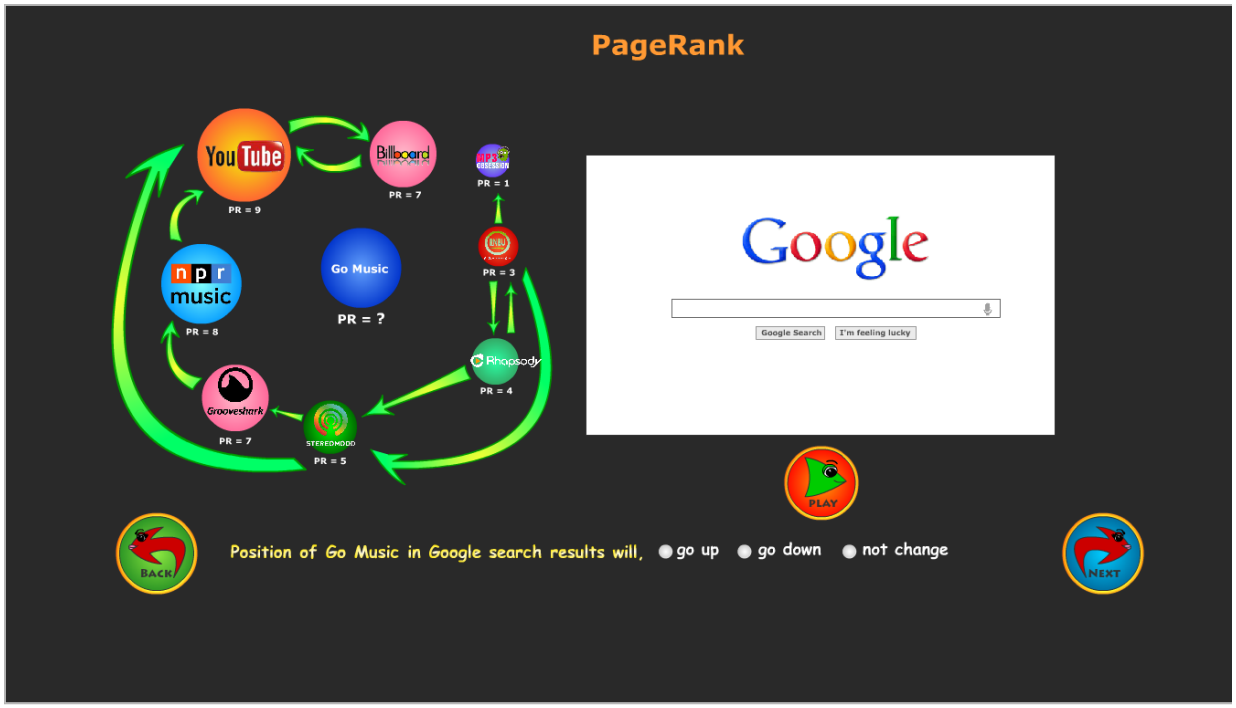

Figure 13 PageRank Interactive Part 3

The initial search results show the position of Go Music web site which is the last. When a node lies around the Go Music is clicked, this position may changed and user has to guess whether the position goes up, down or does not change.

Figure 14 Google Search Results

Chapter 8 – Social Networks

Data Miner explains the idea of Social Networks using Facebook as it is one of the most popular social network site in these days. The main purpose of Social Networks module is to describe certain facts that would help Facebook to connect people as suggestions. With this tool we discussed how Facebook suggests friends to its users based on groups that they have joined, pages they have liked and posts which have been published by them.

Social Networks module also contains a Demonstration and interactive part. The Demonstration explains the idea of degrees of connection by using a network of friends in a social network.

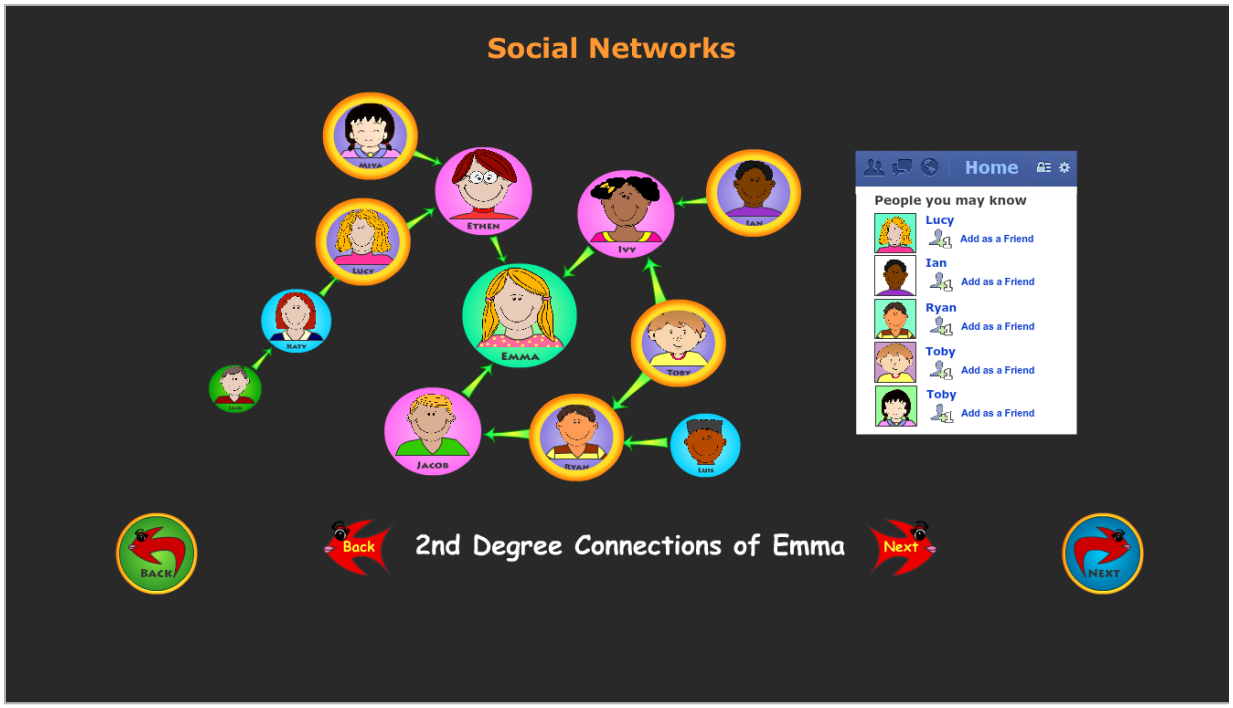

Figure 25 Social Network Demonstration

The Interactive part consists of three levels such that Beginner's level, Intermediate level and Advanced level. In each level, the tool provides some Facebook users with some of their information and asks to suggest a friend to the given user based on certain similarities exist between theses users.

8.1. Beginner's Level

The Beginner's level provides four groups of Facebook users such that each group of users are from either USA, Germany, India or China and a group contains three users. Each user in these groups lives in different cities of their country. The objective of this level is to suggest a friend to the given Facebook user by the system by picking one of these users. If the answer is incorrect, the tool will indicate that and if it's correct then an explanation of the answer will be provided.

Figure 15 Social Networks Beginner's Level

8.2. Intermediate Level

Intermediate level is similar to the beginner's level, but it contains more details about each user than the beginner's level.

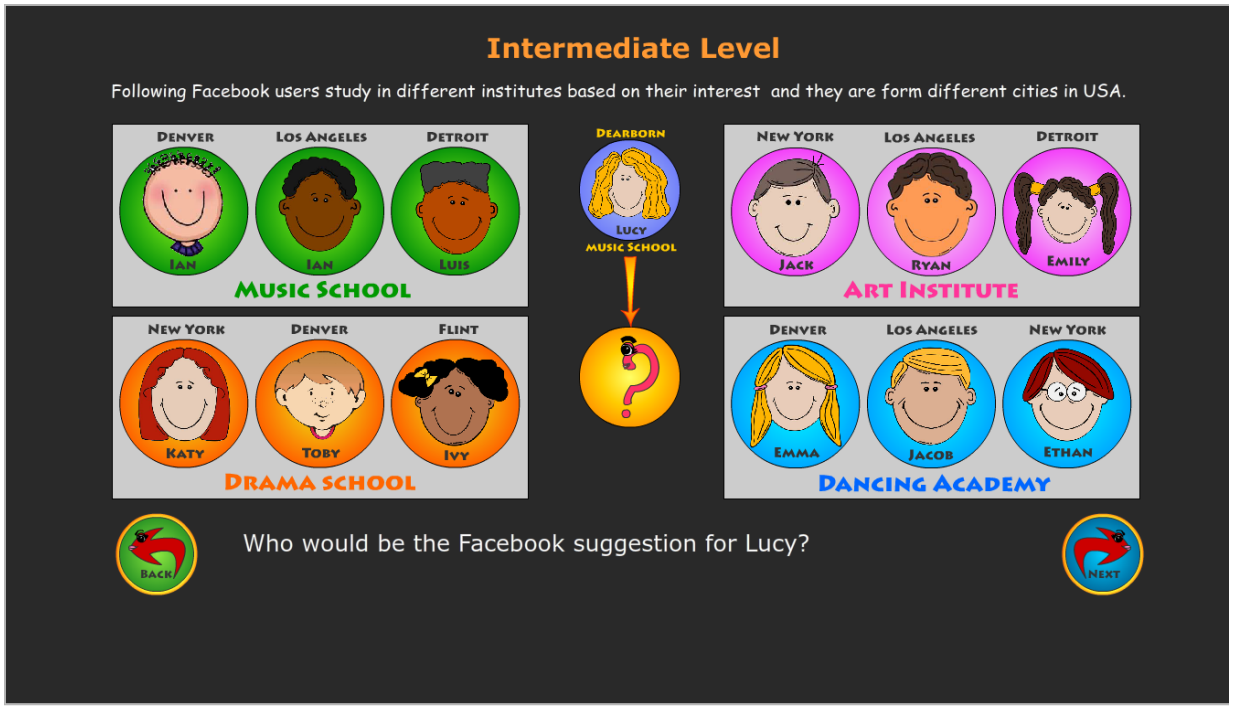

Figure 16 Social Networks Intermediate Level

8.3. Advanced Level

In advanced level the system provided four users with their Facebook Home pages and depending on the similarities of the content in these pages, user has to select the person who can be suggested as a friend to the given Facebook user.

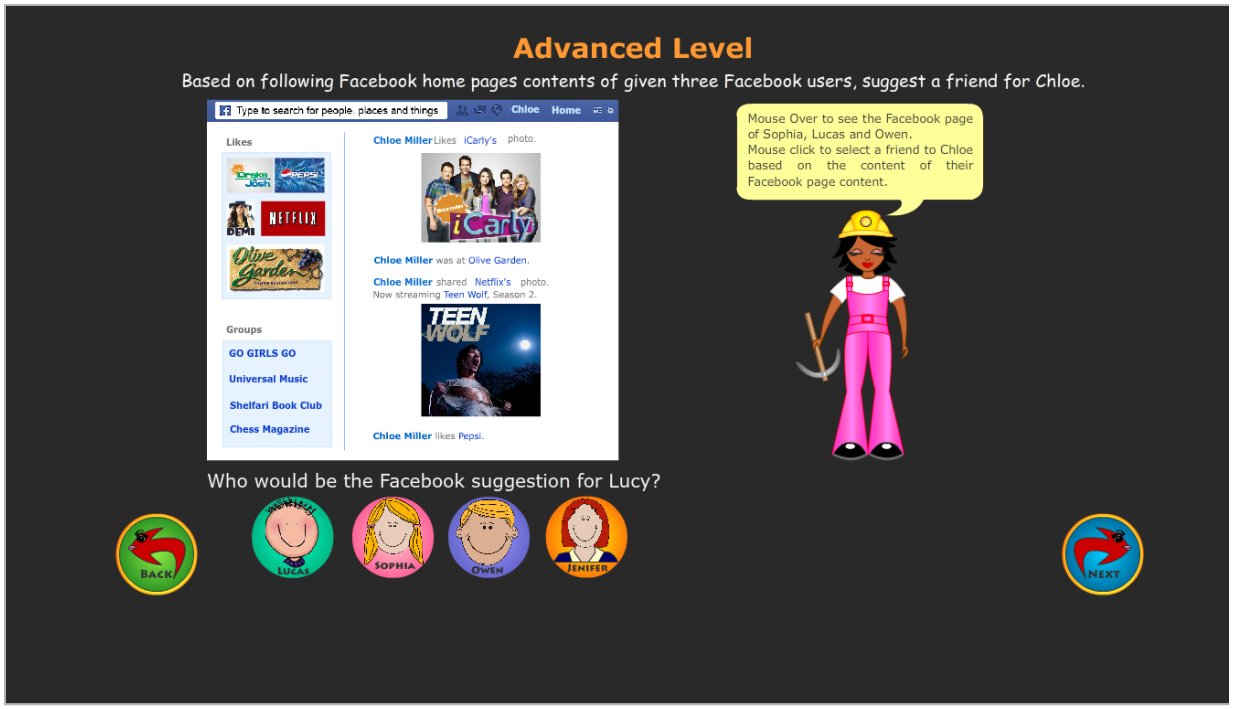
Advanced Level
Based on following Facebook home pages contents of given three Facebook users, suggest a friend for Chloe.
Type to search for people, places and things
Chloe
Home
Likes
Chloe Miller Likes iCarly's photo.
Chloe Miller was at Olive Garden.
Chloe Miller shared Netflix's photo.
Now streaming Teen Wolf, Season 2.
Chloe Miller likes Pepsi.
Groups
GO GIRLS GO
Universal Music
Shelfari Book Club
Chess Magazine
Mouse Over to see the Facebook page of Sophia, Lucas and Owen.
Mouse click to select a friend to Chloe based on the content of their Facebook page content.
Who would be the Facebook suggestion for Lucy?
BACK
LUCAS
SOPHIA
OWEN
JENIFER
NEXT

Figure 17 Social Networks Advanced Level

Initially the Facebook Home pages are not visible, but they will pop up when mouse over event occurs on each button provided as multiple choices.

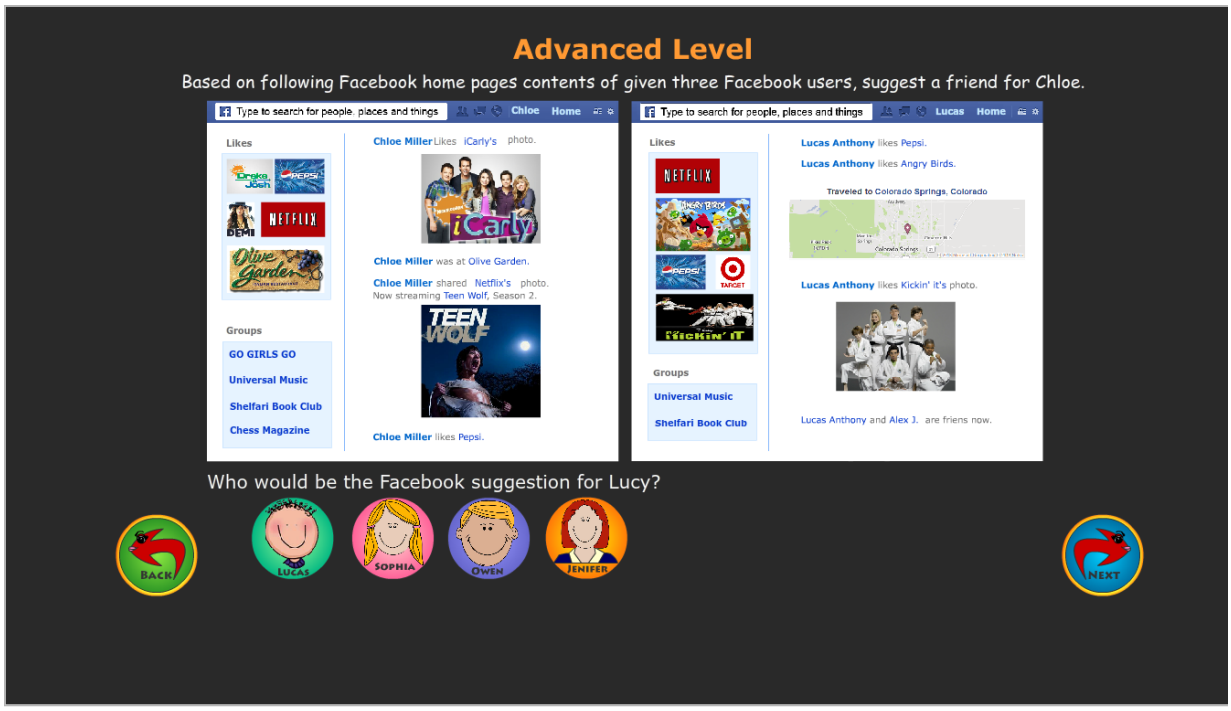

Figure 29 Social Networks Advanced Level Mouse Over on Buttons

Contact Us

Dr. CHANDAN REDDY
Assistant Professor
Department of Computer Science
Wayne State University
Detroit, MI - 48202.

Contact information:
Tel : (313) 577 9005
Fax : (313) 577 6868
Email: reddy@cs.wayne.edu